浅析大数据在图书出版中的运用

摘 要：在科学日新月异的今天，大数据成为时代变革中的主要标志。其善于定位自身发展方向，且作为新时代主流技术之一，大数据与出版业的衔接板块如今产生大量信息处理矛盾，而出版社在大数据洪流之下，则呈现出颠覆式进步，并在借鉴国外大数据辅助图书出版行业基础上，对国内未来出版战略重新进行规划和调整。文章为进一步发挥大数据自身现有优势，围绕国内图书出版行业实际情况对其现阶段所采用出版运营模式进行全面解读，并站在行业高度提出相关建议，以期为出版社未来数字化、信息化发展提供参考价值。

关键词：大数据；图书出版；应用
中图分类号：G239.2　　文献标识码：A
文章编号：1671-0134（2018）08-115-02　　DOI：10.19483/j.cnki.11-4653/n.2018.08.047

文 / 范春龙

1. 应用案例

1.1 大数据与图书出版的有效融合

自 2009 年起，新一代大数据以其较强的融合性成为该行业领域的关键载体，并且影响范围已扩大至图书出版行业。随着大数据时代的到来，部分传统出版行业由于缺乏对新鲜事物的全面认识陷入迷茫。而在短时间内最先适应年轻技术的企业为谷歌，顺应“大数据”趋势，谷歌将“词频统计器”建立在 Google Books（全球图书）庞大数据库之上，形成一种新的应用系统，全球 1500 册藏书内的词汇被纳入其中，网络用户可在网页最多输入 5 个不同的词语，就可以通过 Google Books Ngram Viewer（全球书籍词频统计器）显示的图表观察这些词汇在不同年代数据里的出现频率。与小数据时代不同，大数据时代打破了传统以少量数据信息便可收获一系列信息资源的固化思维模式，其强调对数据本身的观察与研究。

1.2 大数据应用于图书出版领域的弊端

随着电子图书的全面普及，在某种程度上给传统图书出版社带来巨大的压力和冲击，及时采取压低价格等相关手段，仍无法抵御互联网的“吞噬”。大数据的形成与发展，不仅满足广大读者的丰富的阅读需求，且为读者带来新颖化和个性化的阅读体验。从历史角度来看，在人类文明演变过程中出版业担任重大使命，而随着思想潮流的不断变迁，传统出版行业面临来自不同方面的挤压与挑战，例如，数字出版、在线阅读、知识分享等在线娱乐方式。为改变这一现状，国内一批电商企业围绕大数据出版进行了一系列尝试。但国内当当网 CEO 表示：“出版社互联网 + 和图书大数据相结合，未来出版社的价值可能为零。”并列举多项应用弊端，具体如下：

一是面对大量出版社及网络运营商的质疑，当当网 CEO 表示未来利用互联网平台将作者和编辑直接假借，出版社的价值将会为零。例如今日头条，作者写完不需要编辑预审来决定内容的走向，而是直接让读者发表评论，筛出推荐给读者的书。经过当当读书 APP 大数据的统计，每天产生评论 4500 万条。

二是读者在网上购买图书后，可在其他网络社区与其他感兴趣读者分享，虽然实现图书版权的共享，但版权保护防线同时被打破。

1.3 大数据背景下的图书出版行业未来

对于数字出版产业而言，数字化在其平稳运行中起到关键作用，且在大数据时代背景下，数字化已成为大势所趋，全球范围内的数据都有实现共享的可能，而这一目标的实现在某种程度上取决于互联网技术的发展水平。近年来，当当网、亚马逊、京东、新华在线、淘宝等线上图书销售规模呈现明显增长趋势，在做大整体市场规模的同时，对线下书店的分流作用明显。因此，需加快实现图书出版行业与互联网技术的有效融合，充分运用互联网技术，分析所获取的共享数据及单独读取数据等。

2. 在图书出版社中大数据应用需注意事项

2.1 适当采用大数据进行图书策划

大数据技术在图书出版策划中的应用，在条件上有着极为严格的要求。由于当前社会背景下，数据挖掘可为选题策划提供的有益价值相对较少，这一弊端在部分专业极强的图书领域表现尤为凸显，并且差距不断被拉开。其中一部分图书出版社，受自身人员配置及资金等限制，将大数据处理技术应用于“小数据”的处理任务中，以取得良好的图书策划效果。

从某种意义上来讲，大数据技术在我国图书出版社

中的应用，并未达到真正意义上的“大数据”标准，大数据下所产生的数据信息在数量上存在局限。究其原因，主要是由于国内图书出版领域内部缺乏系统完善的大数据治理模式，各出版仍然掌握整个出版行业内的数据获取、互换、反复利用以及存储等工作。不同出版社在人力资源配置、数据投入以及重视程度方面存在一定的差异性，这种行业内的鸿沟成为阻碍图书出版领域数据一体化的主要屏障。国家相关单位针对我国当前图书出版行业现状，对大数据在行业内具体实施标准进行了系统的研究与制定，早在 2013 年便提出出版业数据交换国家标准——《中国出版物在线信息交换图书产品信息格式规范》（CNONIX），以全面实现出版行业内大数据采集、分析、储存及应用等任务的不断升级。尽管该标准的全面实施为图书出版业大数据整体环境的改善提供了有利条件，但仍未在真正意义上建立数据服务平台、数据分析及数据应用等相关产品及相关服务的标准。基于数据买卖和合作的数据挖掘外包服务等乱象重重，尚处于试水阶段，难以在短时间之内进行大范围的推广及应用。

2.2 重点营销图书选题策划建议采用外包服务

当前阶段，大数据方法在出版社选题策划中的应用尚存在技术方面的限制。作为出版选题策划中大数据方法数据的主要来源，内容、读者以及平台三者的绝大部分数据仍掌握在各分社之间，并未有出版社真正掌控。随着数字出版社块度发展，其强大优势日渐凸显，在其发展背后很大一部分产值来自于传统出版内容的快速数据化。但由于大数据在专业性方面的要求较高，且面临庞大的工作量，出版社在当前人力资源条件下，难以按照相关标准完成相关任务，因此便出现频繁的外包现象。与此同时，在对网络媒体以及新媒体中被抓取相关内容进行分析时，往往在技术方面存在较高要求，出版者在自身所培养专业人才以及服务外包两者中进行选择时，普遍偏向于服务外包。销售商掌握着绝大多数读者的阅读渠道，而各大社交网络平台以及阅读平台，则被广大新媒体公司所掌握，出版者能够掌握的数据基础较少，无法形成大数据挖掘所需的大数据库。尽管其中一部分图书出版社将大数据方法有效融入具体的选题策划中，但在实际数据挖掘过程中仍以外包为主。因此，一种以数据买卖及合作的数据挖掘外包服务便由此产生，对于出版单位而言应进一步优化对其应用水平及效率。

2.3 把图书正能量导向作用放在选题策划之前

在部分专业化图书出版社中存在，大数据技术在实际应用过程中在某种程度上存在失真现象，单从技术性角度来看，图书出版专业领域内整体数字化水平仍处于较低阶段。但是与其他产品及服务不同，图书商品在实践中有着明显的特殊性，应将图书商品社会价值的超前性纳入考虑范围之内，即图书社会价值对读者需求向真善美的引领性质。例如，被列入“高冷”行列的科技类图书，通过数十年的实践应用，极有可能引领人们走向另外一条成功道路；人文类图书更是如此，文学类著作往往要“源于生活而高于生活”才能引起人们的共鸣。对于图书而言，对其内容质量的判定并不能一味地寻求与作者共同喜好，更多的是了解掌握广大作者实际阅读需求，并引导读者积极向上、适应社会主流价值观的基础上策划出版相应的图书。此外，大数据在图书出版中的具体应用，应做好与图书相关的阅读内容，所谓大数据在其中仅为理论性引导方法，应尽可能地避免沦为大数据技术的“奴隶”。

结语

综上所述，大数据为图书出版行业带来了新的历史发展机遇，新视野、新方法以及新工具的出现，意味着我国数字化图书出版社的构建应把握大数据驱动下的研究特点与研究规律，掌握最新技术，紧跟信息化时代潮流，不断推进现代图书出版领域理论以及研究方法的创新。因此，在运用大数据思维理念和有形无形的研究方法来回答和解决国内出版业发展所面临的重大理论和实践问题，还应立足先进科学技术前沿，熟练推进大数据、人工智能等信息技术与人文社会科学研究的深度融合，并针对当前大数据在图书出版应用研究中的关键问题，提出中国学者的原创性理论贡献，为数字化图书出版社的构建奠定坚实的基础。媒

参考文献

[1] 张志谦，唐晓丽，王晓祖 . 大数据时代下历史类图书市场走向分析及策划建议 [J]. 中国编辑，2015，16（2）：272−273.

[2] 苗晋诚，文晓利 . 大数据背景下图书出版面临的机遇与挑战 [J]. 传播与版权，2014，1（9）：356−358.

[3] 徐剑 . 大数据时代图书选题信息的采集模式研究 [D]. 安徽大学，2015.

[4] 雷鸣，张钰婷，张大岭，等 . 大数据在学术图书按需出版中的运用 [J]. 出版发行研究，2015，8（11）：281−282.

（作者单位：江西科学技术出版社）